which such forecasts yield, add the number of cases of rain when the pressure is below 29.75 to the number of cases it did not rain when the pressure is above this figure, and divide by the total number of observations.

In this case we get $198+2{,}263 \div 3{,}606 = 0.68$. By this simple rule the forecasts are correct 68 per cent of the time for rain in 12 hours, and by a similar calculation 59 per cent for rain in 24 hours, forecasting rain in the latter case when the pressure is below 29.95 inches. Besson uses this method for the observations at Montsouris with six elements, pressure, direction and velocity of wind, temperature, cloudiness, and pressure change, and obtains percentages of verification ranging from 58 to 67. Traces and days when rain was falling at observation are included in Table 5. Excluding these the winter probability at Dubuque never rises to 0.50 except with northeast and east winds in combination with falling pressure as shown in Table 6. Under these conditions the verification is 73 per cent.

Owing to the low probability of precipitation of measurable amount, the complete application of the above method to the data of this paper is not practicable, and we must revert to the graphical representation of the facts for the essential relationships. The results obtained by Besson and those obtainable from the graphs fall short of those obtained by the experienced forecaster using the synoptic chart, and it is not to be assumed that conditions observed at a single station can give a complete basis for formulating a forecast, as, indeed, the chart and the local observations combined can not do. On the other hand, many weather proverbs and many of the predictions of mariners, farmers, and other outdoor men have a certain validity, due to the fact that locally observable phenomena do precede and announce coming weather changes. A statistical study of local data, by classifying these phenomena in relation to subsequent weather, furnishes valuable supplementary information and suggestion to the forecaster.

551.501

A NEW PRINCIPLE IN THE ANALYSIS OF PERIODICITIES

By CHARLES F. MARVIN

[Weather Bureau, Washington, March 24, 1924]

What appears to be an important new principle has come out in the course of a recent application of the Fourier analyses made by the writer to evaluate a supposed sun-spot period in terrestrial temperatures. This has to do with what happens in the use of the analyses or other schemes for investigating periodicities when the data under examination have within them certain known or unknown periodicities the length of which is shorter than the last term of the Fourier series.

The demonstration which follows is built up around the Fourier series merely as a convenience contributing clearness and comprehensiveness of presentation. The principle itself is concerned only with the body of data investigated and the significance of certain features thereof, and the Fourier theorem is nothing whatever but a symbolism employed to conveniently exhibit the facts.

The present consideration presupposes, of course, that the whole number of phase values of data available is finite and limited, and consequently a finite and limited, although possibly a large number of terms of the series will nevertheless completely reproduce the original observations.

The point brought out in this note may be old and well known, but the writer has not as yet seen it discussed,[1] and experience indicates that it is neither avoided nor adequately discounted in many serious studies of alleged periodicities in meteorological data.

The principle may be stated in the following form:

The several terms of a limited and finite Fourier series do not necessarily represent single or unique harmonic components of the data analyzed, but each term may theoretically at least represent two or even several wholly independent components with widely different periods.

The analysis of this question does not appear to present material difficulties.

Suppose we have a long record, say of temperature for a period of 50 to 100 years or more, and that the individual phase values are weekly or monthly means. Now we know that this entire series of values can be exactly reproduced by a Fourier series with a finite number of terms.

Let K = the large number of phase values in the original record.

Let p = the relatively small time interval between phase values.

Let l = the length of period in the same time units.

Now we learn from the Fourier theorem that there will be a term in the series whenever the fraction $\frac{Kp}{l}$ equals an integer 1, 2, 3, etc.; that is, all the features of the original long record will be expressed as if they were equivalent to a long series of periodicities of designated wave lengths. It is of no significance whatever in this study whether the component periods are real or false. Some may be real and others false, or all may be real or false.

Assuming simply for convenience that K is exactly divisible by 4 we find that the last term in the Fourier series corresponds to the wave length $\frac{Kp}{l=2p}=\frac{K}{2}$ and the whole series appears thus:

$$\frac{Kp}{l}=1,\ 2,\ 3,\ 4,\ 5,\ \ldots\ldots\ldots\ \frac{K}{2} \quad (1)$$

Now we are never able to make practical use of the theory when p is a relatively small interval of time and K is a very large number running into the hundreds or thousands, as assumed above. In practice we content ourselves with making p a relatively large number, P, and K a correspondingly smaller number, k so that the product $kP = Kp$.

It is perfectly obvious that these simplifying assumptions can not in the slightest way affect any of the characteristic features of the original data or the real existence of all of the terms of the series (1) as representing the original data.

We assume in the foregoing that the k phase values at large intervals P are actual observations at the times in question, and we shall continue this assumption for a moment. In actual problems of this kind it is a customary although a faulty practice to assemble a greater or less number of observations contiguous to the desired phase date into a representative mean *as of that date.*

[1] While this note was in process of publication the March issue of the Proceedings of the Royal Society, series A, vol. 105, No. A 731, was received at our library, and the article, A Difference Periodogram, by C. E. P. Brooks, p. 346, brought to my attention. While Dr. Brooks does not discuss in its generality the principle presented by the writer, nevertheless he makes use of it to evaluate short periodicities in the disguise of periods of much greater length.

The effect of the reduction in the number of phase values is simply to divide the original complete series represented by (1) into a series of groups entirely analogous to spectra of the first, second, and higher orders, resulting in the phenomena of diffraction. This is indicated by writing the original series in the following form:

PERIODICITIES OF THE

$$\frac{Kp}{l}=\frac{kP}{l}=\overbrace{1,2,3,\ldots\frac{k}{2}}^{\text{First order.}};\ \overbrace{\frac{k}{2}+1,2,3,\ldots\ldots k}^{\text{Second order.}};$$

$$\overbrace{k+1,2,3,\ldots\frac{3k}{2.}\ldots\frac{K}{2}}^{\text{Third and higher orders.}} \qquad (2)$$

It is not difficult to show that the 1st and 2nd, including other higher *even* orders stand in the relation of rights and lefts to each other, or they are *complementary* in that the wave lengths run from long to short in the 1st and odd orders, whereas they begin short and run to long waves over the sequence of terms as they occur in the periodicities of the *even* orders. In other words, the actually shorter and shorter periods of the 2nd order are, as it were, reflexed back upon terms of the 1st order series, just as if the actual wave lengths were all longer and longer than the real waves of the 2nd and all other higher *even* orders. Each wave of an even order differs by 180° in phase from its complementary wave of the odd order. Again, comparing all the odd orders with each other they may be called companion orders in that the sequence of terms in both cases begin with long and end with short periodicities.

The very discomfiting thing about this whole matter is that all the higher order of periodicities exhibit themselves in disguise just as if they were *bona fide* terms of the first order, which are the only terms that can be evaluated from the k phase values adopted.

The application of the new principle to investigations of periodicities is simply this: All periodicities in the original data of shorter wave length than the shortest one it is expected will be evaluated by the method of analysis employed tend to remain and reproduce themselves, not, of course, as periodicities of very short wave length, but instead as false or disguised periods of long wave lengths. This is especially the case with short periods which are commensurate with the fundamental interval covered by the observations. This result is both helpful and harmful, helpful because the operation of the principle enables one to evaluate with all necessary accuracy the constants of a *known* short period as a long period parading in disguise in the 1st order periodicities. For example, in the analysis of 100 years of monthly temperature values (1,200 in all) the *annual period* was transformed by this new principle into the fourth and eighth harmonics of lengths of 25 and 12.5 years, respectively. While the use of 5 months' means necessarily operated to reduce the amplitudes of the 6 and 12 months' periods appreciably, nevertheless the phase angles were evaluated with great exactness. Both of these periodicities really belong to the fifth order.

The harmful results come about because we can not know in all cases whether the several components of the first order periodicities are *bona fide* relatively long waves or whether they are short-period waves drawn out or disguised into false long waves.

The force of these assertions may appear more clearly if illustrated by a very simple, concrete example, with diagram and a tabulation showing how the terms of the high-order periodicities dispose themselves over the group of terms constituting the first order of periodicities, which are the only ones which can be evaluated.

Suppose we have 48 fundamental phase values of some complex periodic variable for analysis and resolution. Suppose, also, that with the idea of lessening the labor or otherwise simplifying the work, we decide to use only 12-phase values—that is, we adopt a phase value for every fourth fundamental value. Six harmonic terms of the Fourier series will, therefore, completely reproduce the 12-phase values. However, each one of these six components is a possible composite of four entirely independent periods of widely different wave lengths. The heavy dots in Figure 1 represent 12 possible phase values out of the total 48 fundamental observations. These dots are common both to a long wave with a period of 48 fundamental time units, and to a short period of only $3\frac{9}{13}$ time units. These same dots are common also to two other short periods, not drawn, of $4\frac{5}{11}$ and $2\frac{2}{23}$ time units, all of which, with others, could be present in the original 48 fundamental observations.

The lower portion of Figure 1 shows how the 24 possible harmonic components group themselves when one attempts to simplify the analysis by reducing the 48-phase values to 12, as assumed above, resulting in six harmonic components of four orders of periodicities, namely, first, second, third, and fourth. If the reader will imagine the table cut out and wound into a cylinder, it is easy to see how the projections and indentations at the opposite ends interlock and form a continuous and symmetrical arrangement.

It is hoped the foregoing makes very clear how difficult it is to get at the absolute truth concerning complex periodicities.

However, many may be tempted to say yes, but no one would be foolish enough to base an analysis on 12 isolated phase values such as shown by the dots in Figure 1, but would use large group means instead. The answer is, *the isolated values are the only ones capable in the long run of disclosing the truth.* The group means tend to efface amplitude and displace phase characteristics. Large group means and more or less powerful smoothing formula tend to obscure, not reveal, the real but unknown periodic characteristics of data and must be used and resorted to with great caution, discretion, and understanding.

Of course, large group means are highly useful and trustworthy in dealing with great outstanding, well-understood periodicities, such as the 11-year sun-spot period, the annual and diurnal periods of temperature, etc. But our real quandary and search for periodicities are concerned with uncertain and obscure features with very small amplitude.

The foregoing consideration clearly bring us face-to-face with a serious dilemma in the investigation of periodicities. How is it to be met? Obviously, the reduced number of k phase values adopted for final analysis must be entirely free of any possible unknown short periodicities or more or less systematic fluctuation or residuals of such features. Is this prerequisite satisfied by the common practice of taking certain group means or by the use of certain favorite smoothing formulae?

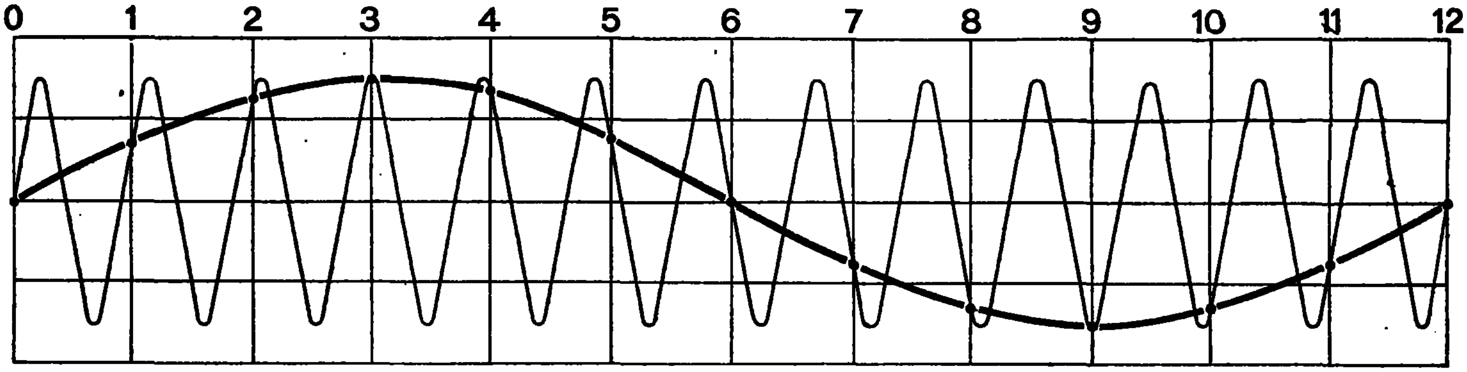
The answer to both these questions is emphatically *no.* A 12-months' mean temperature, for example, is of course free of any possible annual period or exact submultiples thereof, but possible periodic features of 5 months, 7 months, and many other longer wave features are not thus eliminated, and we simply delude ourselves and invite erroneous conclusions and inferences to proceed upon any assumption that such features are eliminated.

In the case of the possible periodicities of long wave length, the only rational course preeminently requires a preliminary investigation to disclose and evaluate all the possible short periodicities or demonstrate that they are absent. If present these short periodicities must be eliminated and segregated before a sane study of the long periodicities is possible. It is not a question whether the short periods are real or false. The preliminary analysis must show what features are present in the data which have the semblance of periodicity. Their existence and presence in the original body of data makes it impossible to draw safe conclusions concerning any long periods, and we must not overlook the fact that so long as the final amplitude of the hidden and obscure periods claimed by some to exist is only of the same order as, or even less than, the size of the residuals which remain uneliminated by the faulty methods of taking group means, thereof must appear in the adjusted data disguised as false long periods. Such disguised long periods are bound to be brought out subsequently if any sort of adequate analysis of periodicities is employed, whether it be the Fourier analysis or any other scheme. At any rate, the writer believes attention has been called to an interesting new principle in periodicities that seems to uncover a kind of pitfall or snare to trap the overconfident claimant of periodicities.

Accidental features and fortuitous fluctuations of data are of great importance in the present connection. These and many other details are fully recognized by the writer but can not be discussed in this brief note.

BASIC PRINCIPLES GOVERNING THE USE OF THE PERIODICITY TABULATION

It seems appropriate to add these brief notes in order to fairly round out and complete this article dealing with the investigation of periodicities.



	1	2	3	4	5	6
12	11	10	9	8	7	
	13	14	15	16	17	18
24	23	22	21	20	19	

Fig. 1.—Representing (above) how 12 phase values out of 48 possible independent observations can represent either a long wave of the first order of periodicities or a short wave of the third order of periodicities, and (below) a tabulation showing the distribution of 4 orders of periodicities of 6 terms each which arise when 48 observed values are reduced by combination or otherwise to 12 phase values.

of smoothing, etc., while these two magnitudes are of the same order, the proof of periodicities is unacceptable or inadequate unless the causes of error herein discussed are avoided.

It seems necessary to repeat here that it is of no consequence that the present demonstration of a new principle in periodicities is built up around the Fourier analysis. The latter is used simply as a convenient symbolism to facilitate a clear and comprehensive presentation. The principle itself really has nothing to do with the Fourier theorem, but declares what happens in any kind of analysis of assumed periodicities when data are arbitrarily smoothed and combined into larger or smaller group means with the declared object of *eliminating* recognized troublesome short periods and irregularities. It can not be claimed that a particular group mean eliminates all possible short periods at the same time. Residuals

94018—24——2

It is well known the familiar device here called the periodicity tabulation comprises r rows of K columns of homogeneous values of data at unit intervals of time and supposed to possess periodic characteristics. A tabulation may be indicated as follows:

$$\begin{array}{lllll}
 & a_1 & a_2 & a_3 \text{-----------} & a_k \\
 & b_1 & b_2 & b_3 \text{-----------} & b_k \\
 & & * \; * & * \; * \; * \; * & \\
 & r_1 & r_2 & r_3 \text{-----------} & r_k \\
\hline
\text{Sums} \text{---------} & S_1 & S_2 & S_3 \text{-----------} & S_k \\
\text{Means} \text{--------} & m_1 & m_2 & m_3 \text{-----------} & m_k
\end{array}$$

The unit of time may for convenience be assumed to be the time interval between columns, as days, months, years, etc. We may then designate the K time units embraced by a single row as the *fundamental interval.*

Similarly, the quantity Kr in time units may be designated the *total interval*. For example, a 4-row 27-column tabulation of *monthly* means would have a fundamental interval of 27 months and a total interval of $4 \times 27 = 108$ months.

Elemental tabulations.—Almost every one thinks of the scheme called a periodicity tabulation as a device to eliminate various features of data, except one supposed elemental period which the tabulation will segregate from all other periodic, nonperiodic, irregularly recurrent and accidental values, and thus evaluate the periodic term alone and by itself. This idea works out beautifully in the case of such outstanding periods as the annual and diurnal periods of pressure, and temperature and annual periods of precipitation and other elements of meteorological data.

Complex tabulations.—However, the tabulation has a vastly wider utility than this. It is just as proper to use it in a manner that will *eliminate* some one particular period like the diurnal or the annual periodicity, and *preserve and develop* numerous other possible periods, if any of them exist. This calls for the use of a tabulation which frequently will have relatively only a few rows but relatively a large number of columns leading to a very complex sequence of final sums and means.

Tabulations of this kind afford by far the most trustworthy evidence for the analysis of periodicities, although experience and judgment are required for their sound interpretation.

Interpretation of results.—The results of every periodicity tabulation must be interpreted under the guidance of important basic principles which are stated briefly in what follows.

Classes of periodicities.—There appears to be no basic distinction between the various possible more or less hidden periodicities in meteorological and like fluctuating data. Nevertheless, purely for convenience in the present analysis we propose to recognize three classes:

1. *Full commensurate periods* are those which are commensurate with the *fundamental interval*.[1] These are

$$\frac{K}{l} = 1, 2, 3 \ldots\ldots\ldots\ldots \frac{K}{2} \qquad (1)$$

2. *Semicommensurate periods* are those which are *not* commensurate with the fundamental interval but which *are* commensurate with the *total interval*. These are

$$\frac{Kr}{l} = 1, 2, 3, 4 \ldots\ldots\ldots\ldots \frac{Kr}{2} \qquad (2)$$

The last terms in (1) and (2) are

$$\frac{K-1}{2} \text{ and } \frac{Kr-1}{2},$$

when K and Kr are odd.

It will be noticed by equation (1) that the possible number of full commensurate periods is fixed and unchangeable as long as the number of the columns K remains constant. In contrast to this, the possible *semicommensurate* periods steadily change and increase as the number of rows r increases.

3. Finally, *full incommensurate* periods are those which can not satisfy the conditions in either class (1) or (2). some of these, however, may, fall into class (2) when a change is made in r.

Rule I.—Every full commensurate periodicity is preserved and developed in all its characteristics by any tabulation, and at the same time the accidental errors and irregular and fortuitous fluctuations, including nonperiodic features present in all data are gradually eliminated, according as the number of rows, r, is increased. There are certain exceptions or reservations arising under this rule which will be mentioned later. (See Rule IV.)

Rule II.—Every semicommensurate periodicity will be *completely eliminated* by a tabulation, *provided* its wave form is symmetrical and the number of rows is sufficient to fairly exclude large accidental features. Even though the wave form of a given period may tend to become symmetrical in the long run, this prerequisite to complete elimination is not satisfied by individual waves, and since certain reasons may sometimes compel the investigator to employ only a limited number of rows, it must be recognized that some residuals from certain semicommensurate periodicities will always remain in the final results. For example, the wave form of the annual period of temperature is well known to be unsymmetrical, hence its perfect elimination because it may be a semicommensurate period is not assured. In fact, the lack of symmetry will be preserved and developed as a short period of some kind.

Rule III.—The sums and means of every tabulation contain residuals of greater or less magnitude remaining from full incommensurate periods. Such residuals tend to be eliminated along with accidental errors according as r becomes large.

Rule IV.—It must be recognized that the apparent length of successive individual periods always suffers change due to accidental causes. Physical causes may also explain even greater changes in apparent length. In such cases, for present purposes we are concerned only with the *average* length of a changeable period over the total interval Kr. If such a period is *full commensurate*, the periodicity tabulation will fully preserve and develop it if the fluctuations in length do not exceed, say, ± 5 or ± 10 per cent of the average length, except that the amplitude will be diminished accompanied by some distortion of the real length of the period all more or less according to the range of variation.

If the fluctuations are numerous and of the order of ± 25 or ± 50 per cent, the whole feature will be quickly eliminated or reduced to small residuals which are indistinguishable from many other residuals and accidental errors always present and due to other causes.

If its average length places one of these changeable periods in class (2) or (3) its fluctuation in length, even when small, combines with the consequences of incommensurability to quickly and effectually eliminate it from the results of a tabulation.

While the author recognizes the full possibilities under the foregoing assumptions, nevertheless, based on any evidence thus far adduced, he is compelled to consider it a misuse of terminology to apply the term "period" or "cycle" to erratic recurrent features which exhibit rapid and frequent changes in length of 50 or 100 per cent, and he must disclaim any belief that such features are due to real physical causes rather than being the product of spurious or accidental recurrences, at least until some reasonable physical explanation thereof is convincingly presented.

[1] Periods which are within a few per cent of *full commensurate* will, of course, quite fully emerge from a tabulation, especially if the number of rows is relatively small, or, better stated, if the excess or deficit of commensurability as a percentage *times the number of rows* is less than about 25 per cent. If sufficient data are available to permit of 3 or 4 consecutive groups of tabulations, the true length of the period can be approximated from the amount by which the phase of the wave is found to advance or retrograde.

Of course, if the law controlling the changes of length becomes known quantitatively it will always be easy to adjust the tabulation to meet such a prerequisite.

Rule V.—As immediate consequences of the foregoing rules the results of any periodicity tabulation will be a composite aggregate of possibly several real commensurate periods, probably of small amplitude, upon which will be superposed numerous accidental errors and many uneliminated residuals, greatly obscuring the actual facts.

In the examination of the results it must constantly be recognized that sequences of wholly fortuitous numbers will always exhibit periodic features, and these, as well as all real commensurate periods, can be evaluated by the Fourier analysis or other devices. Nevertheless, only those features can be claimed as real which emerge and persist and endure in a more or less consistent fashion, regardless of some particular method of derivation. The results must be derived in as many legitimate and different ways as possible. Only those features which consistently survive and emerge from every analysis can be regarded as real periodic features in any body of data. All those features which vanish, change, and reappear incident to every legitimate change of data, method of treatment, etc., must be regarded as quite spurious, unreal, and largely the vagaries of fortuitous conditions.

The whole atmosphere can not, of course, be expected to act as a unit with respect to periodicities, and we must be prepared to find wide differences at different times and in different localities.

If it were not quite foreign to the scope and purpose of this note, it would be most interesting and instructive to show at this point the practical working of the periodicity tabulations on actual data, and the application of the rules in the interpretation of complicated results which are secured. These must, however, be reserved for another time.

551.590.2 : 551.501

FITTING STRAIGHT LINES TO DATA GREATLY SIMPLIFIED WITH APPLICATIONS TO SUN-SPOT EPOCHS

By CHARLES F. MARVIN

[Weather Bureau, Washington, March 24, 1924]

Many studies of the data of meteorology, economics, business, etc., are facilitated and definite results may be expressed by the evaluation of a straight line of best fit to the statistics involved. This is often accomplished in an approximate way by graphical methods, but in a great many cases a far more certain and accurate result can be secured by a very simple arithmetical calculation following rigorously the principle of least squares. Moreover, the computation really entails much less time and effort than that required to produce the less accurate scale drawing of the necessary chart.

The cases in which this simple method can be used arise whenever the data correspond to exactly equal and uniform intervals of time, like days, weeks, months, seasons, etc. In still other cases the observed values correspond to a series of abstract integers like 0, 1, 2, 3, etc., which represent recurrences of certain features, such, for example, as consecutive observations of the epochs or dates of the minima, or maxima of the sun-spot period. Finally, even when the original layout of the statistics does not satisfy the above simplifying condition it is often possible to make some simple adjustments of the data so that the simplifying condition is satisfied. It seems from the foregoing that there are a large number of problems in which the simple computations can be employed, and every student of statistics should be perfectly familiar with it.

The problem is to compute the best values (as defined by least squares) of the constants a and b in the general equation of the straight line,

$$y = a + bx$$

where y represents any series of observations corresponding to integral values of $x = 0, 1, 2, 3$, etc.

In order to accomplish two objects in this same note, I will ask the reader to turn his thoughts for a moment to Newcomb's method [1] of evaluating the normal epochs of sun-spot phenomena and the normal length of the period. His normal value of the sun-spot cycle 11.13 years is widely quoted and universally accepted as probably the best evaluation of this puzzling solar feature. His method must, therefore, be, as it is, a very sound one, nevertheless it seems to be little understood and almost never used, either in the analysis of modern sun-spot data not available to Newcomb or in a hundred other problems of periodicities in other statistical data.

Newcomb's method is simply that of fitting a straight line to the observations which fix the dates of the maxima, the minima, the mid-phase values rising or falling, or any other chosen characteristic of data that may be available, and since the consecutive observed values correspond to successive abstract numbers 0, 1, 2, 3, etc., representing recurrences of the same thing, the simplifying condition of the arithmetical computation is satisfied at the outset.

Both objects of this note, therefore, are accomplished by the calculation of the sun-spot data since, say 1820, to date.

Observations.—We shall use the dates given by Wolfer for simply the minima of sunspots since 1820.

TABLE 1.—*Dates of epochs of minima of sun spots by Wolfer, 1820 to 1924*

x	y	c	x	y	c
0	1,823.3	+3.3	5	1,878.9	+3.9
1	1,833.9	+2.9	6	1,889.6	+3.6
2	1,843.5	+1.5	7	1,901.7	+4.7
3	1,856.0	+3.0	8	1,913.6	+5.6
4	1,867.2	+3.2	[1] 9	1,923.9	+4.9

[1] The epoch of the present sun-spot minimum has not as yet been established accurately, but it will probably differ very little from the date indicated.

Almost every student contents himself with the faulty method of deducing the average length of the period by subtracting the first date from the last one and dividing the difference by 9, viz, $\frac{100.6}{9} = 11.18$ years. This not only presupposes that the first and last dates are exact ones, but it wholly ignores the irregular intermediate dates, and any attempted adjustment of the intermediate epochs to a normal series assigns all the irregularities to the intermediate epochs, while the first and last stand 100 per cent perfect. This is clearly wrong, because we must presuppose that each of the dates is affected by some error or irregularity and determine the amount thereof fairly by the method of analysis. This is what Newcomb's method does.

GRAPHICAL SOLUTION

Procedure.—Lay off on the Y axis of a coordinate diagram a scale of dates beginning preferably a little

[1] Astrophysical Journal, vol. 13, 1901, p. 1.